DOCUMENT RESUME

ED 182 325 TM 010 046

AUTHOR Fruen, Mary
TITLE The Use of Tests in Admissions to Higher Education.
INSTITUTION National Council on Measurement in Education, East Lansing, Mich.
PUB DATE 78
NOTE 9p.
AVAILABLE FROM National Council on Measurement in Education, 1230 17th Street NW, Washington DC 20036 ($1.50)
JOURNAL CIT NCME Measurement in Education; v9 n4 Fall 1978

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Admission Criteria; *College Admission; *College Entrance Examinations; Competitive Selection; Decision Making; *Grade Point Average; Higher Education; Norms; Predictive Ability (Testing); Predictive Validity; *Scores; Standard Error of Measurement; Standardized Tests; Test Bias; Testing Problems; Test Interpretation; Test Reliability; Test Validity
IDENTIFIERS Content Validity

ABSTRACT

There are both strengths and weaknesses of using standardized test scores as a criterion for admission to institutions of higher education. The relative importance of scores is dependent on the institution's degree of selectivity. In general, decision processes and admissions criteria are not well defined. Advantages of test scores include: use of common measure for all applicants; elimination of personal biases; and availability of a common element for comparisons. Disadvantages include: measurement of only one aspect of candidates' abilities; inability to distinguish between comparable candidates; inherent errors in scores; cultural and/or racial test bias; and possible lack of comparability between test content and students' curriculum content. Grade point averages (GPA) are also used as an admission criterion and reflect individual performance over a period of time. Raw scores are usually converted to scale scores, and individual scores are compared to those of a norm group. Factors which must be considered are test reliability, standard error of measurement, predictive validity, and content validity. In conclusion, test scores should be used with other available information about the candidates. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Volume 9, No. 4

NCME measurement in education

Fall, 1978

A SERIES OF SPECIAL REPORTS OF THE NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

ED182325

U S DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NCME

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

The Use of Tests in Admissions to Higher Education*

Dr. Mary Fruen

National Center for Health Services Research

U S Department of Health, Education and Welfare

TM 0 0 0 0 6

ABOUT THIS REPORT

Policies and procedures for admission to higher education have been increasingly scrutinized and legally challenged. Admissions test results are an important factor in selection decisions and users should be well informed about the strengths and weaknesses of tests for this purpose as well as the characteristics of the particular test they are employing. In this article, Dr. Mary Fruen explores a number of key issues concerning the use of tests in admissions to higher education — how admissions tests are used, their pros and cons, technical considerations, and test bias issues.

Dr. Fruen has been involved in both the use and construction of admissions tests. At the University of Toronto Medical School she worked with the Admissions committee in establishing selection policies and proc dures. Subsequently she worked on the New Medical College Admission Test development for the Association of American Medical Colleges. Currently, she is a researcher at the National Center for Health Services Research, HEW, Hyattsville, MD 20782.

PSR

FROM THE EDITOR

My tenure as editor ends with this issue. I should like to thank NCME for permitting me to serve and supporting me when it became necessary. Most particularly let me thank Tom Fitzgibbon whose confidence enabled me to take this position. Herb Rudman begins as Editor of ME with Volume 10. He has already demonstrated that he is a capable, conscientious editor and I am sure he will serve NCME well.

PSR

Standardized tests have become an important tool in the selection of candidates for higher education. Admissions test scores are used, along with information derived from other sources such as the application form, letters of reference, and admissions interview, to select among the applicants for admission. Though admissions tests are widely used, the scores are interpreted in different ways, and the emphasis placed on them in admissions decisions varies. This article explores the ways admissions tests results are used and their strengths and weaknesses for selection decisions. Technical considerations in score interpretation are discussed.

Admissions test scores provide information about the applicant's academic readiness for an educational program supplementing measures of previous academic per-

*The author appreciates the research and editorial assistance of Susan Soucek and the useful comments by Nancy Cole, Sandra Wilson, Janet Cuca, and James Angel.

formance such as grades in high school or college. Admissions tests are generally designed to provide information about an applicant's overall academic aptitude, comprehension of the knowledge base necessary for successful performance in an education program, and/or ability to do the type of reasoning required in the program.

Two major tests are used for college admissions — the Scholastic Aptitude Test (SAT) and the American College Testing program test. The Graduate Record Examination, which has two general aptitude sections and optional sections for several subjects, is used for admission to graduate school. There are generally separate examinations for schools in each of the professions, e.g., New Medical College Admission Test, Law School Admission Test, Dental Aptitude Test.

The major concern in making admissions decisions is to select applicants who can most successfully complete the educational program. Thus, those applicants who appear most likely to succeed on the basis of available information are selected. In addition, particularly for professional schools, it is important to select applicants most likely to become competent practitioners of the profession as well. Committees are often cognizant of their responsibility to supply graduates who meet the varied needs of the public served by the profession. Since a large proportion of professional students graduate, admission to the education program largely determines the composition of the profession.

How Test Scores are Used in Admissions Decisions

The use of test scores in admissions decisions varies considerably, depending on the particular test and the institution and program employing its results in admissions decisions. The test may be a major determinant of the selection of an applicant or it may receive relatively minor emphasis. Decision might be based on a specific numerical standard for test scores and previous academic record or after review of all available information on the applicant. One factor which influences the manner in which tests are used in selection decisions is the degree of selectivity in a particular admissions process.

Wing and Wallach (1971) describe three general categories of colleges according to their degree of selectivity of students: "minimally selective," "moderately selective," and "highly selective." The "minimally selective" colleges have virtually open admissions, and the academic requirements for admission may be simply a high school diploma with no further academic requirements as to achievement in school or tests. Admissions on the open basis may be limited to residents of a particular state or other geographic area. Wing and Wallach do not discuss the use of admissions tests in this type of institution. If used at all in this context, they would probably only eliminate a minimal number of very low scoring applicants from acceptance into the entering class. Such institutions are often actively recruiting students because of declining college enrollments; thus only those students who show insufficient academic promise are rejected on the basis of poor previous academic record or low admission test scores.

"Moderately selective" institutions accept a majority of applicants. Those applicants who are rejected tend to be eliminated on the basis of low scores or inadequate previous academic performance. Though other factors such as special achievements and personal characteristics are emphasized in published admissions policies and might be considered in admissions, Wing and Wallach conclude that these are not stressed in making admissions decisions at "moderately selective" colleges.

"Highly selective" colleges, the majority of which are private, accept only a small percentage of those who apply for admission. The goal in these institutions is to select the best from among many qualified applicants. Stated admissions policies suggest that a goal in selection is diversity, with a multiplicity of characteristics being considered in admissions decisions. However, these insti-

NCME
measurement in education

Volume 9, No. 4 Fall, 1978

Paul S. Redelheim, Editor

Advisory Committee:
Sidney E. Benton, North Georgia College
Carolyn Finley, American Institutes for Research
George S. Ingalls, Portland Public Schools
Anthony J. Polemeni, New York City Board of Education
Warren W. Willingham, Educational Testing Service

Measurement in Education is a series of special reports published four times a year in October, January, March, and May by the National Council on Measurement in Education. These reports are concerned with the practical implications of measurement and related research and their applications to educational problems of individuals, institutions, and systems. The emphasis is upon uses of measurement rather than technical or theoretical issues. Subscription rate: $5.00 per year; single copies 75¢ each in quantities of 25 or more, or $1.50 for a single issue.

Second-class postage paid at East Lansing, MI 48823 and at Cleveland, OH

Address Editorial correspondence to:

Dr. Paul S. Redelheim
George Peabody College for Teachers
Box 158
Nashville, TN 37203

Address Business correspondence to:

Dr. Irvin J. Lehmann
202 S. Kedzie Hall
Michigan State University
East Lansing, MI 48824

tutions seem to place considerable emphasis upon test scores and previous academic achievement.

Admission to most graduate and professional schools such as law and medicine is highly selective. There is a great excess of qualified applicants over places available in the entering class. Thus, admissions are highly competitive, and selection decisions may involve a difficult screening process.

The selection process varies considerably among the three major classes of institutions identified above, as well as among institutions in a particular category and departments or schools within institutions. This process is generally not well-documented. However, it is possible to identify a number of types of admissions processes and look at the role of tests within each.

In the minimally selective and moderately selective admissions processes test results are likely to play a major role, in combination with previous academic record. Test scores and grade point average (GPA) may be considered independently with a cutoff score or minimum acceptable level given for each of these. On the other hand, test scores and GPA may be used in combination, with each weighted according to a mathematical formula to select those applicants who appear acceptable according to these criteria. If there is still a greater number of applicants than available places, further selection from among these may be made either on the basis of the same credentials or through consideration of other available information.

According to Manning (1977), highly selective admissions is often a two-stage process. First, a decision is made on the admissibility of the applicant, that is, his or her qualifications for the educational program in terms of previous education and requisite abilities and aptitudes. Test scores are likely to be a factor in determining acceptability of the applicant at this stage. A cutoff score might be used, with those individuals scoring below a certain level being eliminated from further consideration.

Manning also pointed out that the minimal standards for admissibility must bear a rational relationship to curricular demands, and empirical data demonstrating this relationship should be compiled. He further stresses that such criteria are not infallible. Some applicants who fall below minimal standards might succeed, and others above the minimal standard may do poorly or fail. However, this stage serves to identify applicants who appear to have the best academic qualifications.

The second stage is the selection of a class from among the pool of qualified applicants, generally on the basis of a much broader range of information. The goal is to select those individuals who would comprise the best entering class considering institutional goals and, in the case of the professions, societal needs. Test scores and previous academic performance may be relied on heavily in making the final selection decisions in highly selective institutions. On the other hand, once an individual is determined academically acceptable other factors such as special achievements, interests, personality, and biographic or demographic factors such as place of origin or race may play an important role in the final decision.

Additional sources of information include academic transcripts, test score reports, application forms, and letters of evaluation (Cuca, 1977). In admission to many medical schools and some other professional schools and private colleges, an interview with a faculty member or other representative may also be an important factor in admission decisions. Decision might be made by an individual or by an admissions committee. However, the decision process and criteria actually used at this stage are somewhat elusive since the published reports on this subject are rather meager.

Advantages of the Use of Test Scores in Admissions Decisions

The primary advantage of using admissions tests in selection decisions is that they provide common measures of achievement for all applicants. Students from diverse personal and educational backgrounds can be compared on results of the same test.

Admissions tests are generally standardized. That is, the same test is administered to each candidate under specified conditions. An attempt is made to have testing conditions as nearly uniform as possible. An admissions test for higher education is administered on the same day at testing centers across the country. All test administrators are given identical instructions supplied by the testing service. Each examinee receives the same or equivalent test materials, and all must complete the examination within the same period. These efforts to create uniformity in the testing process minimize the possibility that an applicant's performance will be affected by external factors related to test administration.

Content of an admission test is tightly controlled. Generally, each applicant taking an examination on a particular date receives the same form of the test. However, results are also comparable across test administration dates, because equivalent forms of the test are developed. Questions for each form of the test are written to meet the same content specifications, and test items (questions) are selected to ensure equivalence in breadth, depth, and difficulty. Test length and scoring are also the same across forms, so that a particular score would mean the same thing no matter which form of the test an individual had taken.

Because selection decisions for higher education are so important, development of an admissions test generally involves a number of quality control procedures. Test specifications which define test content are made up, generally by a committee of experts, on the basis of the test purpose and the types of candidates who will be taking it ("Where Did You Ever Get Those Questions," 1978). After test items are written and reviewed, they are administered to a group comparable to the usual examinees. On the basis of the actual performance of the group on each item and considering the desired content balance, items are carefully selected to make up the test. Further, experimentation is likely to be done with the total test to ensure test quality and appropriateness for its intended purpose.

Scores are reported in numerical form. A lot of information about an applicant is distilled into a number facilitating comparison of applicants. The total distribution of scores for all individuals taking an exam is generally published, as are other distributions such as sex and race. These distributions or "norms" provide information about the relative standings of a particular examinee.

Standardization of admissions tests gives the resulting scores an element of comparability not manifest in other sources of information commonly used in admissions decisions. The numeric grade point average might superficially appear to be a uniform measure of performance across applicants. However, even if the grading systems are the same, the course programs of two

applicants and the grading standards of their professors are likely to differ. Interpretation of other common sources of information in admissions decisions — letters of evaluation, interviews, and statements by applicants is inherently subjective. Such selection criteria have been found to be less reliable than test scores and GPA in predicting success in a program of higher education (Cuca, 1977).

Tests are a uniform yardstick. Use of the same measures makes it possible for admissions committees to consider applicants from unknown schools. This is advantageous to those from small or little-known institutions who might not otherwise be given fair access in competitive admissions.

Standardized test results are also free of personal biases. That is, a student has no opportunity to enhance a score on the basis of a personal factor such as charm, nor can the score be downgraded because of an occurrence such as a personal conflict with the teacher or disruptive classroom behavior. Objective scoring minimizes the unreliability of scores due to scorer bias.

Disadvantages of Admissions Tests

Standardized tests, including those used in admission to higher education, have a common disadvantage — they provide information about only limited dimensions of an individual's abilities. While test scores are extremely useful in differentiating among applicants in their performance on specific cognitive tasks, test scores alone cannot approach a complete representation of the potential of an individual. They measure certain specific knowledge and some intellectual skills, but do not assess personal characteristics (such as enthusiasm and dedication), special abilities, relevant accomplishments, etc. Even the knowledge and skills assessed are somewhat limited due to such factors as test length and format. Therefore, it is important not to overemphasize the use of test results in admissions decisions.

Tests are sometimes criticized as measuring recall of unimportant factual knowledge. Though this could occur, admissions tests generally purport to go beyond factual recall. Application of knowledge and reasoning or problem solving can be assessed in a multiple choice item; however, this item type does not lend itself, for example, to testing complex problem solving, writing ability, or creativity.

Because test scores are numeric there is potential of abuse in using them in admissions decisions. An individual who receives a score a point or two higher than that of a second person might be thought to have achieved more in the area tested, when in fact the scores should not be distinguished. An examinee's score should be considered only an estimate of the individual's true ability. It can be appropriately used to identify those who perform very well or very poorly on this type of measure, not to make fine discriminations among individuals.

A certain amount of error is inherent in a test score. Uncontrollable factors in testing render it impossible to develop tests so that scores can be given precise interpretation. For example, the test instructions might be unclear to an individual. This is further argument that tests should not be overemphasized in admissions decisions, and it is inappropriate to consider small score differences among applicants as meaningful.

Admissions tests are often criticized as being biased against particular groups of people. Those with significantly different educational or social backgrounds are likely to show differences in performance on a test. Therefore, caution must be exercised in interpreting scores for such individuals. The issue of test bias is discussed further at the end of this paper.

An admissions test might not fairly reflect the knowledge and skill of a student in the area tested. Because admissions tests have the same content nationally, it might not test the material learned in a particular program of study. Thus, the student may be at a disadvantage in responding to a particular set of test items; also, the student may have related special achievements which there is no opportunity to demonstrate in the admissions test.

In constructing admissions tests there are two approaches to establishing test content so that an individual examinee with unusual preparation would not be unfairly handicapped. First, the exam can assess broad scholastic abilities like verbal skills and quantitative reasoning, which is the aim of the Scholastic Aptitude Test, The American College Testing Program exam, and the Law School Admission Test. A second approach is attempting to limit test content to those areas which are needed for the type of program to which admission is sought. Examples are the New Medical College Admission Test science tests, that assess knowledge of the science areas generally considered prerequisite for medical school, and the Graduate Record Exam Advanced Test section in which the candidate takes the exam in the particular subject relevant for the graduate department to which application is made.

Comparative Advantages of Admissions Tests and GPA in Selection

Since test results are used along with grades to assess academic achievement, it is useful to look at the relative advantages of each. Schwartz (1971) points out that the major advantage of admissions tests is that all applicants are evaluated on the same set of criteria; the main advantage of academic records is that they are based on performance over a period of time. The value of GPA for selection in a highly competitive admissions may have diminished somewhat due to grade inflation. Willingham and Breland (1977) compare the usefulness of grades and test scores in admissions decisions as follows:

> Grades give the most direct evidence of capability and motivation for academic achievement, though they incorporate many potential inequities due to wide fluctuations in grading standards. Admissions tests represent a limited type of talent but they reflect very important general abilities and competence developed through the educational process, and evidence indicates that they serve as a useful yardstick that offsets some of the weaknesses of grades.
>
> A very substantial amount of research supports the validity and usefulness of tests and grades for selection purposes. The previous grade average tends to be a somewhat better predictor than admissions tests at the undergraduate level and the opposite is true at the graduate level, but in each case both measures make a useful contribution.

On the basis of data on four major admissions tests*,

*Scholastic Aptitude Test, Graduate Record Examination, Law School Admission Test, Graduate Management Admission Test.

Willingham and Breland further conclude that "the test plus GPA give a better prediction than either one taken alone; that estimation of success in undergraduate school is more accurate than at the graduate and professional level, and that even the best predictors available are far from perfect."

Scoring

When an admissions test is scored, first the raw score or number of correct items is calculated. Next, the raw score is converted into the "scale score." Scale scores are initially established by examining a group and using the data based on their performance to establish a scale against which the performance of subsequent groups is measured.

Test scores are generally equated from one form of the test or administration to the next. That is, scores are converted to the score scale for the particular test in such a way that scores on different forms of the same test mean the same thing in terms of degree of content mastery. By the statistical process of equating, any differences in difficulty between forms of the test are compensated for when the raw scores or number of items correct is converted into a scale score.

The Scholastic Aptitude Test (SAT), developed for use in decisions regarding admission to college, has a score scale with scores ranging from 200 (low) to 800 (high) with the mean for the initial group arbitrarily set at 500 and the standard deviation set at 100 *(SAT Guide for High Schools and Colleges 1976-77)*. A score of 700 on the SAT Verbal Ability Test, for example, indicates an equivalent performance no matter which form of the test is taken or when it is taken. However, a particular group of examinees might perform differently on the SAT Verbal Ability Test — e.g., the College Bound Seniors of 1974-75 averaged 434 on that test, lower than the mean of 500 for the initial examinee group.

Norms

A test score by itself, in the absence of interpretative data, is essentially meaningless. Therefore, interpretation of an individual's test score is accomplished by comparison with the performance of some reference group. Norm tables, which are included in published manuals for test interpretation, provide the statistical distribution of the performance on the test by particular group(s). Norm tables generally are published for the group of examinees who took the test nationally on a particular date, and often separate distributions are given for subgroups such as sex, minority status, or geographic area. The individual examinee can thus be compared with others in the norm group. Because a national norm group may not be the most appropriate reference group, it might be advantageous to construct a norm table for all applicants to a particular institution or program.

In a norm table test scores are usually listed in descending order, with the percentage who scored below that score level (percentile rank). For example, Table 1, which presents data on SAT Verbal Ability Section for the population of high school seniors who were college bound in 1974-75, lists SAT Verbal Ability scores and corresponding percentile ranks. The percentile rank indicates the percentage of students in the particular group who scored below a certain score. For example, since the score of 550 on the SAT Verbal Ability Test corresponds to a percentile rank of 84, this means that 84 percent of the particular group of students achieved scores of less than 550.

TABLE 1
Data on SAT Scores for College Bound Seniors* in 1974-75

Verbal	
Score	Percentile Rank
800	99+
750	99+
700	99
650	97
600	92
550	84
500	72
450	56
400	38
350	22
300	10
250	3
200	

Mean	434
Standard Deviation	109
Standard Error of Measurement	32
Reliability Coefficient	.90

*Includes the most recent SAT Verbal score for students who were college bound seniors in 1974-75; the most recent test score is included whether the test was taken in the sophomore, junior or senior year of high school.
Source: *SAT Guide*, 1976

In summary, a specific test score on a standardized admissions test generally means the same level of accomplishment no matter which form of the test is taken. However, the percentile rank shows the percentage of a particular group of people who scored below the score level for a particular administration of the test.

Technical Considerations in Score Interpretation

In using test scores in making admissions decisions, it is important to interpret the scores correctly. Two questions are commonly addressed for which answers can be provided in part through research. First, is the test providing a consistent assessment of what is being measured — is the test reliable? Secondly, what are the correct interpretations and uses of the test scores — is the test valid?

Reliability and Standard Error

Uncontrollable factors exist in educational testing, making it impossible to develop tests with absolute precision. Thus, an individual's score should be considered an estimate of his true ability, rather than an infallible measure. Reliability is an indication of the consistency of test results. If a test were perfectly reliable, an individual repeating the same test, or an equivalent one, would receive a similar score (assuming that practice makes no difference).

Since all tests are imperfect, they are therefore unreliable to a degree. It is possible for irrevelant or chance factors to affect the scores received on a particular test. For example, an examinee might be sick, there may be distractions in the testing room, or the test might have unclear directions.

The extent of error in the test score can be estimated. Thus, we can estimate the test's reliability as well. Test reliability is expressed in terms of a correlation coefficient, ranging from 0 to a perfect reliability of 1.00. (Reliability of the SAT Verbal Ability Test is high at .90.)

Because numeric test scores may imply a level of precision which is inappropriate, the potential exists for over-reliance on them. Thus, scores on standardized admissions tests are generally reported along with the reliability estimate and standard error of measurement.

The standard error of measurement is a statistically derived index of the variability to be expected in the scores for a particular test due to measurement error. The "standard error" indicates the probable extent of discrepancy between an individual's obtained score and his hypothetical "true score," which he would receive if the test were completely free of error. Thus, the obtained score is an approximation of an individual's true score. An interval extending one standard error (of measurement) above and one standard error below a particular obtained score would include the true scores of approximately two-thirds of the students who earned that obtained score. Conversely, the true scores of one-third of candidates would fall outside that range. Similarly, the interval encompassing two standard errors would include true scores of roughly 95 percent of candidates.

For example, the Scholastic Aptitude Test (SAT) Verbal Ability Test has a standard error of 32 (approximately 30 in round numbers). Thus, if a student's true score were 450, there are about two chances in three that his actual or reported score will lie between 420 and 480; there is a 95 percent probability that the reported score would lie between 390 and 510 (SAT Guide 1976-77). In interpreting test scores it is wise to construct such a range of scores as a reminder of the level of error in an individual's score.

Validity

Validity is defined in the Standards for Educational and Psychological Tests (APA, 1974) as the "appropriateness of inferences from test scores." Valid tests report scores which faithfully represent a specified domain of skill or knowledge or indicate certain relationships between the scores and other relevant variables. The validity of a test for a particular purpose can be facilitated through specific test development procedures and assessed using statistical and logical analyses. Validity can be defined in terms of a number of different dimensions. Of most concern in validating an admissions test are predictive validity and content validity.

Predictive Validation

The primary objective in construction of an admissions test is that it predicts success in the program to which the candidate is applying. Such prediction is often determined through a longitudinal research study. This involves the analysis of relationships between scores on the admissions measure by a particular examinee group and diverse performance criteria, as they progress through the education program, and possibly into relevant employment.

However, such a longitudinal study is problematic in many respects. One must wait a long time for results. Criterion measures, like admissions measures, have weaknesses in characteristics such as reliability and limitations in the range of knowledge and skills assessed. Methodological problems are numerous including difficulties of attrition of part of the sample, inability to define intervening variables such as training and isolate their affects and interaction among variables. Beyond issues of methodology, one questions the applicability of longitudinal research results in a rapidly changing sociopolitical condition and educational system. For that reason, a second type of predictive validity study — concurrent validation — is needed as well.

In a concurrent validation study both the admissions measure and a criterion measure are administered to the same group within a short period of time. Generally the group included in the study consists of people who are already participating in the educational program for which the admissions test is being developed. The comparison of performance on the admission measure to various measures of achievement in the educational program provides some understanding as to how tests can be expected to function as predictors.

Most large scale admissions testing programs do predictive validity studies. Results are published, often in a manual for test users which provides guidelines for appropriate test use.

Content Validation

Content validity is best established by appropriate test development procedures in order to ensure that the content of an admissions test is most relevant for applicants to a particular type of program and that it reflects as closely as possible the achievements and skills required of individuals entering that type of program. Generally, test items are written according to test specifications which outline the content areas to be included in the examination. This assures that content is balanced across topics considered appropriate for the purpose. For example, in constructing an admissions test for a graduate program in political science, one would attempt to write the content level generally expected of students entering such a graduate program.

Issues in Validation of Admissions Tests

Criteria for admission to higher education and employee selection have been increasingly questioned and challenged. Much of the controversy focuses on the issue of validity. Manning (1978) summarizes the conclusions on validation approaches reached in both the Federal Guidelines on Employee Selection Procedures and in court cases. Though these primarily involve employee selection, the use of tests for that purpose is somewhat analogous to college admissions. Manning points out that in test validation, emphasis has traditionally been on predictive studies, both long-term and concurrent. The guidelines of the Equal Employment Opportunity Commission (1970) stressed the importance of predictive

validity studies. However, interpretation of what constitutes valid use of test results has broadened and shifted to stress the importance of content validity. That is, instead of strict empirical studies, content validation is based on logical analysis of the extent to which knowledge or skills tested are reasonable bases in selection for a particular job or higher education program.

This shift in emphasis from predictive validity to content validity is evident in the outcome of a number of major court cases. In 1976, the Supreme Court refused to review a decision by the Court of Appeals in Tyler vs. Wickery, a case concerning the use and validation of the Georgia bar exam. One part of the decision required that the examination test skills and knowledge with a "logical, apparent relationship to those necessary for the job." In the case of Washington vs. Davis in 1976, the Supreme Court ruled that a verbal ability test could appropriately be used in the selection of candidates for police training. This decision suggests that not only could an admissions test legitimately assess general abilities required for success in the educational program, but also that the reasonableness of content was in itself sufficient validation evidence without the necessity of a predictive study.

Are Admissions Tests Fair?

Critics have questioned the fairness of standardized tests, claiming that such tests are culturally biased and therefore discriminatory. The debate over the fairness of tests is ongoing, and the social and political issues raised in this discussion are numerous and complex. Increased awareness of this issue has been further stimulated by recent legal action as well.

An important issue in the use of tests in admissions is the question of fairness to all applicants. The process of making admissions decisions should be fair; that is, information on each applicant should be used appropriately in the selection decision.

"Bias" of an admissions test can also refer to test content. Content should be balanced and demonstrably appropriate to the purpose for which it is used. Items that are unusually difficult for a particular group should be carefully examined for possible sources of difficulty irrevelant to the purpose of test use.

A related consideration is that the admissions test predict the success of the individual in the educational program. It should add data to the information which will help to select those from the applicant pool who are likely to have the best potential as students. The preponderance of results from prediction studies with the major admissions tests shows that test scores do typically predict educational success for a variety of groups.

A fundamental problem which arises in any discussion of fairness or cultural bias in the use of standardized tests is that no consensus exists concerning what constitutes fairness or equality of opportunity. Linn (1973) in a review of the literature discussed several alternatives of test bias and implications of each for candidate selection. The technical problems and contradictory values in each model were pointed out. Flaugher (1978) concluded that what one judges as fair test use in selection depends ultimately on values, as no selection procedure is totally objective.

A well constructed test measures inequality where it exists; it does not create this inequality. The standardized test is designed to differentiate among individuals along

REPORTS AVAILABLE

Back issues of *Measurement in Education* are available for 1.50 each per single issue, and for 75¢ each in quantities of 25 or more for a single issue.

Vol. 1,	No. 1	*Helping Teachers Use Tests* by Robert L. Thorndike
	No. 2	*Interpreting Achievement Profiles — Uses and Warnings* by Eric F. Gardner
	No. 3	*Mastery Learning and Mastery Testing* by Samuel T. Mayo
	No. 4	*On Reporting Test Results to Community Groups* by Alden W. Badal & Edwin P. Larsen
Vol. 2,	No. 1	*National Assessment Says* by Frank B. Womer
	No. 2	*The PLAN System for Individualizing Education* by John C. Flanagan
	No. 3	*Measurement Aspects of Performance Contracting* by Richard E. Schutz
	No. 4	*The History of Grading Practices* by Louise Wittner Cureton
Vol. 3,	No. 1	*Using Your Achievement Test Score Reports* by Edwin Gary Joselyn & Jack C. Merwin
	No. 2	*An Item Analysis Service for Teachers* by Willard G. Warrington
	No. 3	*On the Reliability of Ratings of Essay Examinations* by William E. Coffman
	No. 4	*Criterion-Referenced Testing in the Classroom* by Peter W. Airasian and George F. Madaus
Vol. 4,	No. 1	*Goals and Objectives in Planning and Evaluation: A Second Generation* by Victor W. Doherty and Walter E. Hathaway
	No. 2	*Career Maturity* by John O. Crites
	No. 3	*Assessing Educational Achievement in the Affective Domain* by Ralph W. Tyler
	No. 4	*The National Test-Equating Study in Reading (The Anchor Test Study)* by Richard M. Jaeger
Vol. 5,	No. 1	*The Tangled Web* by Fred F. Harcleroad
	No. 2	*A Moratorium? What Kind?* by William E. Coffman
	No. 3	*Evaluators, Educators, and the Publics: A Detente?* by William A. Mehrens
	No. 4	*Shall We Get Rid of Grades?* by Robert L. Ebel
Vol. 6,	No. 1	*On Evaluating a Project: Some Practical Suggestions* by John W. Wick
	No. 2	*Measuring Reading Achievement: A Case for Criterion — Referenced Testing and Accountability* by S. Jay Samuels and Glenace E. Edwall
	No. 3	*Security in a Citywide Testing Program* by Anthony J. Polemeni
	No. 4	*Dear Mama: Why don't they love me anymore?* by Thomas J. Fitzgibbon
Vol. 7,	No. 1	*The Discrepancy Evaluation Model (Part I)* by Andres Steinmetz
	No. 2	*The Discrepancy Evaluation Model (Part II)* by Andres Steinmetz
	No. 3	*What Graduate and Professional School Students Think About Admission Tests* by Letnard L. Baird
	No. 4	*The Paradox of Education Testing* by Robert L. Ebel
Vol. 8,	No. 1	*FACE UP: A Framework for Assessing Career Educators Utilization Programs* by Bruce W. Tuckman
	No. 2	*Grade Equivalent Scores* by Gary Echternacht and *If Not Grade Equivalent Scores — Then What?* by Jeanne M. Plas
	No. 3	*Criticisms of Testing: How Mean is the Median?* by Henry S. Dyer
	No. 4	*The Superintendent and Testing: Implications for the Curriculum* by Herbert C. Rudman
Vol. 9,	No. 1	*A Consumer's Guide to Criterion — Referenced Test Item Statistics* by Ronald C. Berk
	No. 2	*Perspective on Intelligence Testing*, by Roger T. Lennon
	No. 3	*Bias*, by Dr. Eric F. Gardner

certain dimensions (such as level of knowledge in a particular subject area). Thus, to the extent that differences in background among groups exist in the types of behavior tested, one would expect differences in test results. However, one would not want the test to differentiate among individuals along irrevelant dimensions. For this reason, test publishers typically submit tests to thorough reviews by various minority reviewers. To avoid bias the test must be carefully constructed and its use validated. In addition, users should consider the validity in a specific context for a particular type of individual.

It is well established that there are differences in performance on standardized tests which are correlated with race, socioeconomic status, and gender. Thus, the use of test results has important social implications. These implications must be weighed against other goals in admissions decisions; value judgments are made implicitly or explicity in test use for selection purposes. It is not suggested that the use of tests should be eliminated, but rather that the issue of fairness must be resolved in score interpretation. Caution should be used in attaching meaning to scores and using them in the admissions process.

Summary

Admissions tests are important tools for decisions on selection of any applicants for higher education programs. However, it is important that test results be interpreted carefully, taking into account factors such as validity of their use, nature of the content, and measurement error. Test scores are best used in combination with other available information on the applicant. Though test scores and previous grades have been shown to be the best predictors of success in programs of higher education, other student characteristics relevant to later performance are appropriately considered as well.

Standardized admissions tests generally have an interpretative manual with suggestions as to their use. For advice on interpretation of the results of a particular admissions test, careful study of this manual is recommended.

Cuca, Janet M., Sakakenny, Linda A., and Johnson, Davis G., *The Medical School Admissions Process*, Washington, D.C.: Association of American Medical Colleges, 1976.

Flaugher, Ronald L., "The Many Definitions of Test Bias," *American Psychologist, 33*: 671-679, 1978.

Linn, Robert L., "Fair Test Use in Selection," *Review of Educational Research, 43*: 139-161, 1973.

Manning, Winton H., "The Pursuit of Fairness in Admissions," *Selective Admissions in Higher Education*, San Francisco: Jossey-Bass, 1977.

Manning, Winton H., "Test Validation and EEOC Requirements: Where We Stand," *Personnel*, 1978.

Scholastic Aptitude Test Guide for High Schools and Colleges, Princeton, N.J.: College Entrance Examination Board, 1976.

Schwartz, Paul A. "Prediction Instruments for Educational Outcomes," *Educational Measurement*, Robert L. Thorndike, Editor, Washington, D.C., American Council on Education, 1971.

Standards for Educational and Psychological Tests, Washington, D.C.: American Psychological Association, 1974.

"Where Did You Ever Get Those Questions?" *ETS Developments*, Princeton: Educational Testing Service, Fall, 1978.

Willingham, Warren W., and Breland, Hunter M., "The Status of Selective Admissions," *Selective Admissions in Higher Education*, San Francisco: Jossey-Bass, 1977.

Wing, Cliff W. and Wallach, Michael A., *College Admissions and the Psychology of Talent*, New York: Holt, Rinehart and Winston, Inc., 1971.

NCME NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

202 S. Kedzie Hall
Michigan State University
East Lansing, Michigan 48824

Second class postage paid at East Lansing, Mich. 48823 and additional mailing points.

MELVILLE S D
EDUCL TSTG SERV
PRINCETON NJ 08540